THE OBSERVATIONAL APPROACH TO COSMOLOGY

Messier 51

THE OBSERVATIONAL APPROACH TO COSMOLOGY

BY

EDWIN HUBBLE

OF THE MOUNT WILSON OBSERVATORY
CARNEGIE INSTITUTION OF
WASHINGTON

OXFORD
AT THE CLARENDON PRESS
1937

OXFORD UNIVERSITY PRESS
AMEN HOUSE, E.C. 4
London Edinburgh Glasgow New York
Toronto Melbourne Capetown Bombay
Calcutta Madras
HUMPHREY MILFORD
PUBLISHER TO THE UNIVERSITY

PREFACE

THIS book contains the Rhodes Memorial Lectures delivered at Oxford in the Autumn of 1936, under the general title, 'The Observational Approach to Cosmology'.

The observable region of space, the region that can be explored with existing instruments, is a sample of the universe. If the sample is fair, its observed characteristics should furnish important information concerning the universe at large. The lectures describe the general features now known, and discuss the nature of the inferences to which they lead.

The features, however, include the phenomena of red-shifts whose significance is still uncertain. Alternative interpretations are possible, and, while they introduce only minor differences in the picture of the observable region, they lead to totally different conceptions of the universe itself. One conception, at the moment, seems less plausible than the other, but this dubious world, the expanding universe of relativistic cosmology, is derived from the more likely of the two interpretations of red-shifts. Thus the discussion ends in a dilemma, and the resolution must await improved observations or improved theory or both.

However, the significance of the investigation lies not in the failure to reach a unique solution to the problem of the structure of the universe, but rather in the fact that the venture is now permissible. As late as fifteen years ago the observable region was restricted to our own system of stars, the system of the Milky Way. Since that time great reflectors have identified the nebulae as independent stellar systems, the true inhabitants of space. Explorations, using the nebulae as gigantic landmarks, have swept out beyond the Milky

Way to the very limits of existing telescopes. The observable region, our sample of the universe, has been suddenly magnified a million million fold. Now, for the first time, the sample may be fair.

The break through into extra-galactic space and the preliminary reconnaissance of the observable region have been described in *The Realm of the Nebulae*, recently published by the Yale University Press. The Rhodes Memorial Lectures form a sequel to the story, for they present the results of accurate surveys which followed the reconnaissance and suggest their cosmological significance. Since the new results could not be discussed in complete isolation it has been necessary to include a considerable background derived from the earlier investigations. It is a pleasure to acknowledge the courtesy of the Yale University Press in permitting the generous use of material from *The Realm of the Nebulae*.

Although the subject is developed from the observers' point of view, it is necessarily permeated with cosmological theory. Fortunately, the writer has had the privilege of association with Richard C. Tolman of the California Institute of Technology, who has presented the theory in a manner especially adapted to the limitations of the observational technique. Any errors in the application of the theory must be attributed to the misuse of his friendly counsel.

The illustrations reproduce photographs made with the telescope chiefly responsible for the recent development of the field of nebular research, namely, the 100-inch reflector of the Mount Wilson Observatory of the Carnegie Institution of Washington.

E. H.

MOUNT WILSON OBSERVATORY
CARNEGIE INSTITUTION OF WASHINGTON

CONTENTS

LIST OF ILLUSTRATIONS

PLATES

DIAGRAMS

CHAPTER I

THE OBSERVABLE REGION AS A SAMPLE OF THE UNIVERSE

The Observational Approach to Cosmology

THIS series of lectures concerns the observational approach to cosmology, to the study of the physical universe. From our home on the earth we look out into the dim distance, back into the dim past, and we strive to imagine the sort of world into which we are born. Observations now range through an immense volume of space: perhaps the nature of the universe may be inferred from the appearance of the sample we explore. Theory presents us with an infinite array of possible universes, logically consistent systems: perhaps our information is now sufficient to identify among them the particular type, or family of types, which includes the actual universe we inhabit. At any rate, astronomy has developed to the point where, for the first time, such attempts are justifiable. Empirical investigations have definitely entered the field of cosmology. Already, certain conclusions can be drawn from the explorations. The long process of elimination and successive approximation has begun.

Cosmology lay for ages in the realm of sheer speculation. Rational arguments were introduced slowly until the critical period just two decades ago. Then theory invaded the field in force, and rapidly exploited the possibilities offered by general relativity. Later still, a dozen years ago, observations crossed the frontiers of the stellar system, and swept out into the universe at large. The observable region of space, our sample of the universe, is now defined, and a preliminary reconnaissance has been completed.

Very recently, the reconnaissance has been followed by accurate surveys that extend out to the practical limits of the largest telescope in operation. Further significant extensions are not expected until more powerful methods are developed, or still greater telescopes are constructed. Therefore, the time is appropriate for a provisional interpretation of the data already available. The conclusions are tentative but they are none the less impressive, for once again, as in the days of Copernicus, we seem to face a choice between a finite, small-scale universe and a universe indefinitely large plus a new principle of nature.

The Copernican Revolution

The universe of the Greeks was a small sphere. The centre was the earth; the boundary, the thin shell of the fixed stars. Between, lay the orbits of the sun, the moon, and the planets. The whole heavens, it was believed, rotated once a day around the motionless earth. The daily rotation of the heavens, carrying along all the contents—sun, moon, planets, and stars—was directly responsible both for the small scale of the universe and for the thin boundary shell of stars. The larger the universe the more tremendous would be the speed of rotation. Deliberate efforts were made to keep the universe as small as possible. The orbits of the planets were packed together as closely as their supposed motions would permit; the stars were confined to a thin boundary shell, pressed tightly around the orbit of the outermost planet. The minimum radius eventually adopted was 20,000 times the radius of the earth, 80 millions of miles. Even so, a star on the celestial equator travelled 6,000 miles per second in its daily rotation.

The Greek conception of a small, closed universe was accepted almost without question until the Coper-

nican revolution. Copernicus not only transferred the centre of the planetary system from the earth to the sun but also transferred the daily rotation from the heavens to the earth. No longer did the huge universe whirl madly about the motionless earth; instead, the earth itself, at a modest rate, was spinning in a motionless universe.

Thus Copernicus removed the necessity for a small-scale universe, and for a thin boundary shell of stars as well. He was primarily concerned with the planets and did not himself take the logical step of scattering the stars through infinite space. That step was reserved for an Englishman—Thomas Digges—who, forty-three years later (1576), assumed it as an obvious and welcome consequence of the new cosmology. The restless imagination, it seemed, had been loath to postulate a universal boundary but had been forced to that extremity by what appeared to be the hard facts of nature. Now there emerged a possible, alternative interpretation of those facts. A choice was presented between a small, bounded universe, centred around the observer, and a universe indefinitely large plus an unfamiliar principle of relative motion.

Definite observational evidence to guide the choice was slow to materialize.. Nevertheless, the larger universe, with all its significance, was adopted quite early, on the grounds of simplicity and uniformity. It was recognized that the system of the planets, centred around the sun, was isolated and lonely in space; that the sun itself was but one of the myriads of stars scattered at vast intervals through the universe.

The Theory of Island Universes

The first notions of the scale of stellar distances in the new universe were derived by Newton and by Huygens, using the principle of the uniformity of

nature. If the sun were a star, they argued, it would be reasonable to assume that all the stars were like the sun—in particular that they all had the same intrinsic luminosity, the same candle-power, as the sun. The assumption is now known to be only a very rough approximation to the truth. Nevertheless, it justified the estimation of the general orders of distances from the apparent faintness of the stars. If a star were really as bright as the sun, but appeared a million million times fainter than the sun, then it would be a million times as far away—about 16 light-years. Using this method, Huygens estimated that the distance of Sirius, the brightest star in the sky, was about half a light-year, and Newton estimated that the distance of stars of the first magnitude was about 15 light-years.

These estimates are only a few per cent. of the true distances, but they gave the first intimation of the immense scale on which the stars are scattered. Speculation could now orient its flight, and soon a still larger vision emerged. In 1750 Thomas Wright published the first approximately correct explanation of the Milky Way. The stars, he said, are not scattered indefinitely through the universe; they form a limited system, isolated in space. The stellar system is rather flat, like a disk or a coin, and the sun is near the centre. When we look towards the top or bottom of the disk, the distance to the boundary is short and the eye encounters few stars. When we look towards the rim, the distance to the boundary is great and the eye encounters many stars, which, in projection, appear crowded together to form the Milky Way.

Wright carried his speculations still farther. The notion of a single stellar system, alone in the universe, was unwelcome to his sense of proportion. He dreamed of a universe populated by countless similar stellar systems, separated by vast intervals. As visible evi-

dence he pointed to certain faint cloudy patches in the sky, known as nebulae. These mysterious bodies, he suggested, were the nearest of the neighbouring systems.

Five years later Kant developed these speculations in a form which was immediately accepted, and which persisted unchanged until recent years. Eventually, the conception was called the theory of island universes—stellar systems scattered through the ocean of space.

Observations followed as rapidly as instruments and technique developed, and, step by step, they tended to confirm this particular line of speculation. Sir William Herschel sketched the rough outlines of our stellar system by counting the numbers of stars his telescope revealed in different directions. The relative numbers of stars, he considered, indicated the relative distances to the boundary of the system along the various lines of sight. About a century ago distances of stars were measured by direct triangulation. Then powerful indirect methods of estimating distances were developed, and the realm of the stars could be explored with a measuring rod. Finally, the stellar system was mapped, using the very luminous globular clusters as landmarks. The stars, it was fully demonstrated, do form a definite system, isolated in space. Beyond the boundaries, the universe stretched on and on, inaccessible to actual investigation but populated in fancy by other stellar systems, comparable with our own.

Thus the second great chapter in the exploration of space was developed. The first had been confined to the realm of the planets; the second ranged through the realm of the stars. And now, in our day, the third chapter has opened. For the explorations have at least won their way beyond the stars into the realm of the nebulae.

Nebulae as Island Universes

The break through was an achievement of great telescopes and especially of the greatest of them all, the 100-inch reflector on Mount Wilson. The story is briefly as follows. Nebulae have long been known as faint mysterious patches of light, scattered with the stars over the face of the sky. One, the Andromeda nebula, is readily seen by the naked eye, and the numbers increase rapidly as the luminosities diminish. In the course of time several thousand nebulae have been catalogued individually; to-day, perhaps 200,000 are recorded on photographic plates. The greatest telescope, under the best conditions, records as many nebulae as stars.

Study of the nebulae revealed two quite different types. The one consists of clouds of dust and gas illuminated by neighbouring stars. These objects, numbering a few scores in all, are members of our own stellar system—the galactic system, or system of the Milky Way. They show a decided preference for the plane of the Milky Way—the galactic plane—and for this reason they are known as 'galactic' nebulae (or sometimes as nebulosities). They will not be further discussed in these lectures.

The other type of nebulae consists of the regular, symmetrical bodies, many of them showing a spiral structure, found by the thousands everywhere in the sky except in the Milky Way itself. Positive information concerning their true nature began to accumulate about a quarter of a century ago, and, by 1924, their status was determined. They were demonstrated to be independent stellar systems as the theory of island universes had supposed.

The conspicuous neighbouring systems were so near that, with the 100-inch reflector, many of their brightest

stars could be photographed individually. Among these stars, various types were recognized which are familiar in our own stellar system. They were all super-giants and their intrinsic luminosities were fairly well known. Therefore, their apparent faintness indicated the distances of the nebulae in which they were found. In all cases where the method could be applied the nebulae lay far beyond the boundaries of our own stellar system. They were scattered through extra-galactic space, and, consequently, they have been called 'extra-galactic' nebulae. In these lectures the adjective will be dropped and the systems will be called nebulae.

Once the flood-gates were opened a wave of exploration surged forward. Already there were large accumulations of data which awaited only the essential clue, the scale of nebular distances, for their interpretation. Further accumulations followed, and now they were planned with a true perspective.

The new investigations followed two lines. In the first place, the more conspicuous nebulae were studied individually in order to determine their structure and contents, to discover their common features, and to devise general methods of estimating distances. Then, with the nature of the inhabitants known, and the scale on which they are scattered, the characteristics of the observable region as a whole were examined. The general investigation was made in two steps, a preliminary rapid reconnaissance followed by a series of accurate surveys. These lectures concern the last stage of the investigations but, in order to clarify the significance of the final results, it will be convenient to summarize the reports of the preliminary studies.

Family Characteristics of Nebulae

The nebulae were found to be members of a single homogeneous family. They are all of the same general

order of intrinsic luminosity (or candle-power), and they exhibit a common pattern of rotational symmetry about dominating central nuclei. Although the structural forms vary widely, they fall naturally into an ordered sequence in which the fundamental pattern varies systematically from one end to the other. The sequence begins with compact globular nebulae and proceeds through a series of flattening ellipsoids into a series of unwinding spirals. Many other characteristics, including colour, spectral type, and resolution into stars, also vary systematically through the sequence; consequently, the sequence has been adopted as the basis for the general classification of nebulae.

It is possible that the sequence represents the life-history of nebulae. The observations alone are not decisive, but they follow with remarkable fidelity the course indicated by the theory of evolution developed by Sir James Jeans. If the theory is even approximately correct, we may provisionally assume that the globular nebulae are young and the open spirals are old. Our own stellar system, an open spiral, would then be classed as an old, mature organization.

The life-history of nebulae is a question for future investigation. For the present we are concerned with the sequence as evidence of a family relationship which permits all the nebulae to be reduced to a standard type and discussed as a single homogeneous group. In this way we can describe the average nebula, and determine numerically the scatter of various characteristics around the average values. The scatter is not great; the nebulae are curiously similar.

This fact greatly simplifies the interpretation of data from surveys. The intrinsic luminosities of nebulae are so alike that, where large members are concerned, they can often be treated, for statistical purposes, as though they were precisely constant. The apparent

a. N.G.C. 4501. Messier 88. *b*. N.G.C. 4486. Messier 87

NEBULAE IN THE VIRGO CLUSTER

Orientation, south to the left

A cluster, since it is a sample collection of many nebulae, all at the same distance, furnishes information concerning the dispersion in size, luminosity, and structure among nebulae. The Virgo cluster, at a distance of 7·5 million light-years, is the nearest, and offers the most detailed information. All types of nebulae are found, although the frequency of early types, and especially of elliptical nebulae, is much higher in the cluster than in the general field.

The cluster members reproduced in the plate represent the two ends of the sequence of classification. Messier 87 (N.G.C. 4486) is a globular nebula. It is unique in being the only object of its type which exhibits a trace of resolution: faint stars can be seen in the extreme outer regions of the nebula.

Messier 88 (N.G.C. 4501) is an open spiral of the same general type as Messier 51 (reproduced in the frontispiece).

faintness of the nebulae then indicate their relative distances. The final step, the determination of absolute distances, depends upon the numerical value of the mean intrinsic luminosity. This essential datum could be evaluated only when the mean distance had been derived for a representative sample collection of nebulae.

Criteria of Nebular Distances

Since a reliable scale of distance is of vital importance in the exploration of space, the simple principles on which the scale was established will be discussed at some length. If the intrinsic luminosity (or candle-power) of an object is known, the apparent faintness indicates the distance. Conversely, if the distance is known, the apparent faintness indicates the intrinsic luminosity. On these principles the development of distance criteria proceeded step by step.

Fundamental distances are derived from certain easily recognized types of giant stars whose intrinsic luminosities are well known from investigations within the galactic system. These stars, among which the Cepheid variables are the most important, range from 1,000 to 4,000 times as bright as the sun. They have been identified in several of the neighbouring systems, which, together with the galactic system, form a local group, more or less isolated in the general field. The existence of this local group is a fortunate chance. The members constitute a small sample collection of nebulae whose distances are well determined by familiar, established methods.

A study of the sample collection furnished by Cepheid variables demonstrated that the very brightest stars in the different nebulae are about equally luminous. They average about 50,000 times as bright as the sun, and the individual deviations are small. Thus

the brightest stars furnish a second criterion which indicates the distances of all nebulae in which any stars can be detected, regardless of whether or not the particular types can be recognized. The distances are not individually accurate, but mean values of even small groups should be reliable. The 'resolved' nebulae, about 150 in all, form a second, larger sample collection which is believed to be fairly representative.

Analysis of this second collection leads to a third criterion of distance, namely, the total luminosities of the nebulae themselves. The nebulae average about 1,700 times brighter than their brightest stars—in other words, about 85 million times brighter than the sun. The scatter is rather large. The brightest giants are about ten times brighter than the average, and the faintest dwarfs, about ten times fainter. Nevertheless, the majority fall within the narrow limits from one-half to twice the average of them all. Thus the criterion is essentially statistical. Individual distances are very uncertain, but mean results for large numbers of nebulae are quite reliable.

Total luminosities form the general criterion which, in a statistical sense, applies to all the millions of nebulae that can be recorded with existing telescopes.

Two additional criteria should be mentioned, because, although their application is limited, they furnish individual distances of nebulae in which stars cannot be detected. One is the law of red-shifts which, as will be explained later, are displacements of spectral lines towards the red from their normal positions. Red-shifts are directly proportional to the distances of the nebulae in which they are observed. The law was established with the aid of statistical distances previously available, but, once established, it furnishes reliable individual distances for all nebulae whose spectra can be recorded. The data are important because they

do not depend upon resolution. For instance, the individual distances derived from red-shifts indicate clearly that the different types of nebulae are strictly comparable. There is no appreciable systematic variation in total luminosities in the sequence of nebular types.

Finally, there are the brightest nebulae in clusters. Some twenty or more great clusters of nebulae are known, each containing several hundred individual members. They are remarkably similar. We find large clusters of large bright nebulae, small clusters of small faint nebulae, and tiny clusters of tiny, extremely faint nebulae. They appear as though we were observing a single cluster from appropriately different distances. Thus their relative distances are readily estimated from their apparent characteristics, and the actual distance of any one—for instance, the nearest cluster—would lead directly to the actual distances of them all. The actual distances are important because, among other reasons, each cluster is a large sample collection of nebulae. The score or more of clusters furnishes a sample collection containing several thousand individual nebulae of all types. Analysis of this collection should indicate the absolute characteristics of nebulae in general.

The clusters are so similar that the mean luminosity of, say, the ten brightest members, or even the individual luminosity of, say, the fifth brightest nebulae, forms a precise and convenient measure of distance. The criterion has been calibrated; for instance, the fifth brightest nebulae are about 650 million times as bright as the sun, and the variation among different clusters is almost negligible. Therefore, the brightest nebulae in clusters furnish the criterion which applied over the greatest range in distance. The great clusters are the most remote objects to which precise individual distances can be assigned.

The step-by-step development of distance criteria is

rather impressive. We start with familiar methods, currently used for investigations within the galactic system, and assemble a small sample collection of nebulae. The collection, as a whole, calibrates a second method, individually less precise, but ranging to much greater distances. With the second method we assemble a fairly large sample collection of nebulae. This second collection, as a whole, calibrates a third criterion, still less precise for individual cases, but ranging out to the very limits of the telescope.

The greatest uncertainty lies in the second step. The brightest stars are a certain number of times brighter than the Cepheids in the particular nebulae which form the first sample collection; but the collection is small, and we cannot be sure that it represents a fair sample of nebulae in general. Nevertheless, it is the only sample we have and, since the data are internally consistent, we must make the best of it. Otherwise, the procedure seems to be thoroughly reliable. Resolved nebulae are a certain number of times brighter than their brightest stars, and unresolved nebulae are strictly comparable. Red-shifts and data from clusters emphasize the consistency of the general picture.

DESCRIPTION OF PLATE III

The great clusters are the most remote objects to which individual distances can be assigned with confidence. For this reason they play a leading role in the formulation of the law of red-shifts. The clusters are all very much alike. Each contains several hundred members, among which the elliptical types predominate. The fifth brightest nebula averages about 650 million times as bright as the sun (very nearly the same as the mean of the ten brightest members), and serves as a convenient criterion of distance.

In the Coma cluster the fifth nebula appears about 2,000 times fainter than the faintest naked-eye stars, and, consequently, the distance is about 45 million light-years. The average red-shift in the cluster is $d\lambda/\lambda = 0{\cdot}245$, corresponding with a velocity of recession of about 4,500 miles per second.

The cluster consists of perhaps a thousand nebulae scattered over several square degrees. The plate covers an area of 190 square minutes of arc, and shows about 85 members of the cluster. The most conspicuous nebulae are N.G.C. 4874 (globular) and N.G.C. 4884 (elongated).

THE COMA CLUSTER

The Average Nebula

The scale of distance, as previously mentioned, is the essential clue to the interpretation of explorations in the realm of the nebulae. The significance of surveys, for instance, depends upon the characteristics of the average, or normal, nebula. From a study of the relative frequencies of various types we know that the average nebula is an intermediate type spiral. We know, further, that it is about 85 million times as bright as the sun, and perhaps 1,000 million times as massive. The conspicuous portion, called the main body, is a highly flattened, lens-shaped figure, about 8,000 light-years in diameter. Very faint exterior regions can be traced by delicate photometric methods out to a diameter of about double that of the main body. The nebula is a stellar system whose contents are rather similar to those of the galactic system. The very brightest stars are blue super-giants which are most numerous in the outer regions of the spiral arms. The integrated light from the main body is very similar to sunlight.

Surveys, in general, deal with nebulae so distant that practically all their characteristics are lost from view except their total luminosities. The surviving characteristic assumes unusual importance, and requires further discussion involving a rather technical point. Nebulae in a given volume of space, or those in a great cluster, average about 85 million times as bright as the sun. However, nebulae of a given apparent faintness average about 2·5 times brighter, or about 210 million suns. This curious relation arises from the dispersion in luminosities together with the more or less uniform distribution of nebulae. A group of nebulae which appear equally faint contains giants, normal nebulae, and dwarfs. The giants are more distant than the

dwarfs, and, consequently, are scattered through a larger volume of space. Therefore, the giants greatly outnumber the dwarfs, and the average for the entire group is evidently brighter than normal.

In either case—nebulae in a given volume of space or nebulae of a given apparent faintness—the percentage deviations from the mean luminosity are scattered at random, and the numerical value of the dispersion is known. Therefore, when we assume that all the nebulae have the same average luminosity, we can assign the numerical values of the uncertainties. In the case of a single nebula the chances are about equal that the error is greater than, or less than, 50 per cent. For the mean of groups of nebulae the probable error diminishes as the size of the group increases. The probable error in the mean luminosity of a group of a hundred nebulae, selected at random, is about 5 per cent.

These examples illustrate one of the advantages of statistical methods which deal with large numbers of individual objects. Accidental errors, random deviations from normal, tend to cancel out. The major source of uncertainty in the results arises from the possibility of systematic errors. In the surveys, for instance, the effects of dispersion in luminosities can be accurately calculated; relative distances of nebulae are known rather precisely. Consequently, results which depend upon relative distances only, and which are derived from logarithms of distances, are quite reliable. Uncertainties arise when actual distances are introduced, because the unit of distance may be wrong. The most likely source of error, as previously mentioned, is the step from Cepheids to brightest stars. A careful examination of this and other possible sources of error suggests that the unit of distance is probably correct to within 25 per cent. However, the main re-

sults of the surveys, and those that will be discussed in these lectures, would not be materially altered if the unit of distance were in error by as much as 50 per cent.

Distribution of Nebulae over the Sky

Now let us consider the surveys. The nebulae are great beacons scattered through space. We know something about their nature, and, in particular, we know their intrinsic luminosities. Therefore, we can study their distribution—investigate the characteristics of the observable region as a whole. There are two immediate problems—the distribution over the face of the sky, and the distribution in depth. In both cases the method of investigation depends upon the simple proposition that, in a statistical sense, apparent faintness indicates distance.

The distribution over the sky is examined by comparing the numbers of nebulae per unit area which are brighter than a particular limit of apparent faintness. In principle the method is simple and direct, but faint limits must be selected in order to include large numbers of nebulae. Statistical averages are significant only when they represent large populations.

The results of such surveys indicate that the *apparent* distribution over the sky is not uniform. Nebulae are not found along the heart of the Milky Way, and they are scarce along the borders. Beyond this zone of complete or partial avoidance the numbers of nebulae per unit area increase directly with galactic latitude right up to the poles of the Milky Way.

The departures from uniformity in the apparent distribution follow a familiar pattern, namely, that due to local obscuration, to the absorption of light within the galactic system. From our position within the system we look through the swarm of stars into the universe beyond. But the system is not completely transparent.

Great clouds of dust and gas are scattered among the stars, and they are especially prevalent along the central plane which defines the Milky Way. These clouds, piling up one behind another, completely obscure the very distant stars, and, of course, the more remote nebulae as well. Cloud absorption fully accounts for the zone of nebular avoidance which follows the Milky Way.

In addition to the scattered clouds, the main body of the galactic system seems to be embedded in a very tenuous medium which appears to be fairly uniform. Whatever its actual structure may be, the medium behaves roughly as though it were an extended uniform layer, centred on the galactic plane. From our position near the sun the shortest paths through the 'uniform layer' are perpendicular to the galactic plane, towards the galactic poles. In these two directions the absorption by the medium is least—about 25 per cent.—and the nebulae are most numerous. As the line of sight departs from either pole and approaches the Milky Way, the path through the uniform layer lengthens, the absorption increases, and the numbers of nebulae decrease. This *latitude effect* is similar to the fading of the sun as it drops from the zenith to the horizon and its light travels along a constantly increasing path through the earth's atmosphere.

The apparent distribution over the sky must evidently be corrected for local obscuration before the true distribution is revealed. Absorption by the tenuous uniform layer is readily corrected; the latitude effect follows the familiar cosecant law. Along the belt of the Milky Way, however, the absorption is practically complete, and, consequently, the proper corrections are unknown. The true distribution of nebulae within the zone of avoidance cannot be determined from the observations.

Outside the zone of avoidance, the large-scale, true distribution is thoroughly uniform. The two galactic hemispheres are closely alike, and there are no systematic variations in either latitude or longitude. The uniformity in the explored areas of the sky runs up so smoothly to the edges of the unexplored regions that we have no hesitation in assuming that the uniformity extends over those latter regions as well.

Minor irregularities do exist. Nebulae are found singly, in pairs, triplets, and groups of various sizes up to the occasional great clusters. Our own nebula, the galactic system, is the chief component of a triple system, in which the two Magellanic Clouds are the satellites. This triple system, along with a few neighbouring nebulae, forms a typical, small group, isolated in the general field. In fact, the nebulae exhibit a pronounced tendency towards clustering. However, the tendency seems to operate on a modest scale; for instance, no cluster is known with as many as a thousand members. When large areas of the sky are compared, the irregularities average out, and the large-scale distribution appears to be thoroughly uniform. Our sample of the universe is isotropic, very much the same in all directions.

Distribution of Nebulae in Depth

The large-scale uniformity emerges from all surveys to faint limits, when local obscuration has been corrected and the limits have been carefully determined. Each survey can be specified in numerical terms. To a particular limit of apparent faintness, say to stellar magnitude m, there are, on the average, a certain number of nebulae per square degree, say N_m.

This method of describing the data leads to the second problem of nebular distribution, namely, the distribution in depth. When several surveys are made

to different limits of apparent faintness, each survey, although it represents the examination of many thousands of nebulae and requires months or years for its completion, may be summarized by the single symbol, N_m. The symbol may be interpreted in various, equivalent ways, because the limiting faintness, m, represents a specific distance and a definite volume of space. For instance, we may compare the various surveys in order to determine how the numbers of nebulae increase with the volumes of space they occupy. If the numbers are strictly proportional to the volumes, and $N = V \times$ constant, we know that the distribution of nebulae in depth is uniform. If the factor of proportionality is not constant, we know that the distribution departs from uniformity, and the departures become very significant features of our sample.

The first counts were made hurriedly—a rapid reconnaissance for the purpose of planning the accurate surveys that followed. The result of the surveys will be discussed later, after certain corrections required by red-shifts in nebular spectra have been explained. At the moment we will consider only the preliminary counts. They demonstrated that the large-scale distribution in depth is roughly uniform. Within the uncertainties of the data, the numbers of nebulae were found to be a constant multiple of the volumes of space they occupy. There was no evidence of a thinning out with distance. The observable region, our sample of the universe, is approximately homogeneous as well as isotropic; everywhere and in all directions, it is very much the same.

The Observable Region as a Sample of the Universe

The homogeneity indicated by the reconnaissance, even as a rough approximation, is very significant. The uniform distribution extends out to the limits of our telescopes. There is no trace of a physical boundary,

no evidence of a super-system of nebulae isolated in a larger world. As far as the observations can be interpreted, the realm of the nebulae may be the universe itself, and the observable region may be a fair sample.

This proposition could not be formulated before the reconnaissance was completed. As long as our positive information was restricted to the stellar system alone, the observable region then available could not possibly be regarded as a fair sample. The stellar system was known to be finite and isolated. Beyond the boundaries lay the universe, unknown, but necessarily different from the star-strewn space within the stellar system. But now we have explored a certain portion of that outer space. Our observable region has been suddenly enlarged a million million fold. It is populated with nebulae, uniformly distributed out to the very limits. If the nebulae do form a super-system, it must be so immense that our sample is wholly insignificant. As far as the reconnaissance is concerned, the notion of a super-system is mere speculation, and, as such, is unnecessary and uneconomical. Let us, then, follow the principle of the uniformity of nature and accept the observable region as a fair sample of the universe. The assumption will serve as a reasonable working hypothesis until it leads to contradictions. Then it can be revised or replaced to conform with our new information.

The picture suggested by the reconnaissance is a sphere, centred on the observer, about 1,000 million light-years in diameter, throughout which are scattered about 100 million nebulae. The nebulae average about 85 million times as bright as the sun, their over-all diameters average between 15,000 and 20,000 light-years, and the average separation between neighbours is about 2 million light-years. A suitable model would be furnished by tennis balls, 50 feet apart, scattered through a sphere 5 miles in diameter.

We know further that the average mass of the nebulae is about 1,000 million times the mass of the sun, and, consequently, we can assign a numerical value to the smoothed-out density of nebular material in space. The value, between 10^{-29} and 10^{-30} grammes per cubic centimetre, is evidently a lower limit to the mean density in the observable region because it ignores matter that may lie between the nebulae. The fact that we have not been able to detect any matter in inter-nebular space does not necessarily exclude its existence, even in considerable quantity, but it does suggest that the minimum density, derived from the nebulae alone, is probably not far below the true value.

The important features of the observable region, considered as a sample of the universe, are: first, the approximate homogeneity; secondly, the general order of the mean density; and, thirdly, an additional characteristic that has not yet been described. The third feature, which will be discussed at length in the next lecture, is the law of red-shifts—the observed fact that light from a distant nebula loses energy in proportion to the distance it travels from the nebula to the observer.

The uniform distribution of nebulae and the linear law of red-shifts suggest that our sample of the universe is too small to indicate its nature. When we consider regions beyond the limits of our telescopes, we can only extend these simple features, on and on, indefinitely. The inferences we can draw are negative at best. Positive information concerning the universe can be derived

---

DESCRIPTION OF PLATE IV

The Corona Borealis cluster is almost a replica of the Coma cluster, but, being nearly three times as remote, appears smaller and fainter. The fifth nebula appears about 20,000 times fainter than the faintest naked-eye star, and indicates a distance of about 130 million light-years. The red-shift is $d\lambda/\lambda = 0{\cdot}0707$, corresponding with a velocity of recession of 13,150 miles per second.

The cluster is compact, for the average distance between neighbouring members is less than 100,000 light-years, as compared with about two million light-years between nebulae in the general field. The illustration, centred about 16′ north of the star B.D. $+27°$ 2482, represents an area about 8 per cent. of the area of the full moon, and shows about one hundred nebulae.

THE CORONA BOREALIS CLUSTER

only from systematic variations whose trends are established within the sample. None was found in the preliminary reconnaissance. Perhaps larger telescopes would reveal them, but larger telescopes are not yet available.

There is, however, another way in which possible variations may be sought. The preliminary results were approximations. The distribution of nebulae was found to be uniform, and the law of red-shifts was found to be linear, within the uncertainties of the investigation, but the uncertainties were considerable. Greater precision might reveal systematic departures from uniformity, and from linearity, even within the observable region as defined with existing instruments. The deviations, if they exist, will evidently be small quantities, and will emerge only from extensive and very accurate data, ranging out to the greatest possible distance. The problem is of first importance, and the answer must be sought with the largest telescopes in operation. A considerable body of the necessary data has now been collected with the Mount Wilson reflectors, and the results will be presented in the lectures which follow.

To anticipate, the investigations lead to alternative pictures, depending upon the alternative possible interpretations of red-shifts. If red-shifts are the familiar velocity-shifts, systematic variations do exist in the observable region, and they suggest an expanding universe that is finite, small, and young. On the other hand, if red-shifts are evidence of some unknown principle of nature, which does not involve actual motion, then variations are not appreciable in our sample, and the observable region is an insignificant fraction of the universe as a whole. Thus, in a certain sense, we again face a choice between a small finite universe and a universe indefinitely large plus a new principle of nature.

CHAPTER II

THE ROLE OF THE RED-SHIFTS

THE previous lecture discussed the nature of the nebulae, their intrinsic luminosities and their apparent distribution. The nebulae are stellar systems, more or less comparable with our own system, scattered through space as far as telescopes can reach. Their luminosities vary, but not widely; on the average, they are about 85 million times as bright as the sun, and, consequently, they can be observed and studied in very remote regions of space. Since they are so nearly alike, their apparent faintness measures their distances. Therefore, the apparent distribution can be determined by counting the numbers of nebulae to successive limits of apparent faintness. The extensive data now available indicate that the *apparent* distribution thins out with increasing distance, systematically in all directions.

However, the *apparent* distribution is not the *true* distribution. Distances are estimated on the basis of the simple law of inverse-squares. Of two equally luminous bodies, if one appears four times fainter than the other, it is supposed to be twice as distant; if a hundred times fainter, ten times as distant. The results are precise, provided the apparent luminosities are not influenced by any factors other than distance. But other factors must be considered. Space-absorption, for instance, or rapid motions in the line of sight, might presumably affect the apparent luminosities. The first, space-absorption, is now known to be negligible, but the second presents a question for investigation which, as we shall see, is still unanswered.

One other factor remains to be considered. This factor—red-shifts in nebular spectra—quite definitely affects the apparent luminosities of the nebulae, and in

a very conspicuous manner. The apparent distribution of nebulae and, indeed, all results involving estimates of distances, must first be corrected for effects of red-shifts before the actual map of the observable region can be drawn. Therefore, the present lecture will be devoted to the observed behaviour of red-shifts, and their effects on apparent luminosities. We enter at once into the rather technical field of spectrum analysis, but this is a necessary preliminary for the more general discussion in the concluding lecture.

Spectrum Analysis

Let us begin with a simple description of spectra. Light is radiated in waves of many different lengths. The eye rather crudely distinguishes the different wave-lengths as colours—long waves are red, and short waves are blue or violet. When we look at a luminous body, the eye receives a beam of composite light—many different colours, mixed in different proportions. However, if the beam of light passes through a glass prism, or other suitable device, the individual rays are bent in different degrees, depending on the wave-length, and the colours are spread out in an ordered sequence called a spectrum. The rainbow is the familiar example.

The sequence never varies. From the long waves of the red, the wave-lengths steadily diminish to the short waves of the violet. The spectrum may be long or short, depending on the apparatus, but the relative positions in the sequence remain unchanged. Position in the spectrum indicates the wave-length of the particular light in question; relative brightness at the position indicates the relative abundance of the particular wave-lengths in the composite radiation. Therefore, a spectrum furnishes valuable information concerning a distant light-source because it indicates the particular colours that are radiated, and their relative abundance.

For instance, an incandescent solid, such as an electric light filament, radiates all possible colours; the spectrum is *continuous* from red to violet, and beyond in either direction. The relative abundance of the various colours measures the temperature of the light-source.

Again, an incandescent gas, such as a neon sign, radiates only certain particular colours. The spectrum, known as an *emission* spectrum, is a pattern of isolated colours separated by dark gaps. The pattern is characteristic of the particular gas involved, hence an emission spectrum serves to identify the chemical composition of a distant, inaccessible light-source.

There is still a third type of spectrum, and it is of especial interest for our immediate purpose. When an incandescent solid, or equivalent source, giving a continuous spectrum, is surrounded by a cooler gas—for instance, a star with its surrounding atmosphere—then the gas absorbs from the continuous spectrum of the background just those colours which it would emit if it were itself incandescent and isolated.

The result is called an *absorption* spectrum. It presents a fairly continuous background, interrupted by dark gaps or lines which represent the colours absorbed by the superimposed gas. The distribution of intensities in the continuous spectrum indicates the temperature of the star, and the pattern of dark absorption lines identifies the gases in the stellar atmosphere.

Absorption lines in the spectrum of the sun indicate the presence of many or most of the elements known in the laboratory. Lines due to metal vapours are predominant, the most conspicuous being a close pair of calcium lines in the violet region, known as the *H* and *K* lines. Other conspicuous lines are due to iron and to hydrogen.

Now the spectra of nebulae are very similar to the spectrum of the sun. Stellar systems, it appears, are

dominated by yellow dwarf stars like the sun. The nebulae in general are so faint that their light can be spread only over very short spectra. Nevertheless, even on the small scales, the *H* and *K* lines of calcium are strong and unmistakable, and the more conspicuous lines of iron and hydrogen can also be identified.

Red-Shifts

Although the spectra of the sun and of the nebulae exhibit the same pattern of absorption lines, there is one remarkable difference. The lines in the nebular spectra, in general, are not in their normal positions; they are displaced towards the red end of the spectrum to positions representing wave-lengths somewhat longer than normal. The entire pattern of absorption lines, all details in a spectrum, appear to have been shifted towards the red. These displacements are commonly known as *red-shifts*. They are characteristic features in the spectra of all nebulae except a few that are in the immediate neighbourhood of our own stellar system.

Each line in a given spectrum is shifted by a certain constant fraction of its normal wave-length λ. The linear shifts vary with the wave-lengths, but the fractional shift, $d\lambda/\lambda$, remains constant. Therefore, the red-shift in a particular nebula is specified by $d\lambda/\lambda$, and it is $d\lambda/\lambda$ which varies from nebula to nebula.

The study of many nebulae has shown that, on the average, the red-shifts increase with the apparent faintness of the nebulae in which they are measured. Therefore, we conclude that, on the average, red-shifts increase with distance. Extensive investigations have demonstrated that the relation is approximately linear,

$$\text{red-shifts} = \text{constant} \times \text{distance}$$
$$d\lambda/\lambda = k \times r.$$

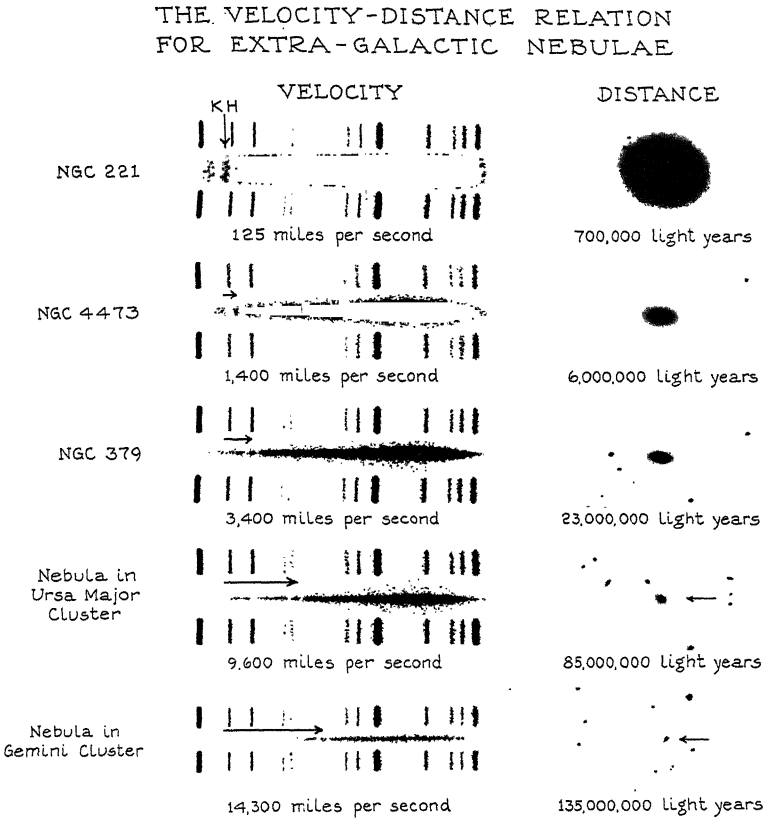
This relation is called the law of red-shifts.

Possible Interpretations of Red-Shifts

When first observed the red-shifts were immediately attributed to radial motion away from the observer, to recession of the nebulae. This interpretation still remains the only permissible explanation that is known. It is true that other ways are known by which red-shifts might be produced, but in each case they would be accompanied by other phenomena which would be conspicuous and, actually, are not found. We may state with some confidence that red-shifts are the familiar velocity-shifts, or else they represent some unrecognized principle of nature. We cannot assume that our knowledge of physical principles is yet complete; nevertheless, we should not replace a known, familiar principle by an *ad hoc* explanation unless we are forced to that step by actual observations.

Most of the theoretical investigators adopt this point of view, and accept without question the interpretation of red-shifts as velocity-shifts. They are fully justified in their position until evidence to the contrary is forthcoming. But these lectures will present a remarkable situation. The familiar interpretation of red-shifts seems to imply a strange and dubious universe, very young and very small. On the other hand, the plausible and, in a sense, familiar conception of a universe extending indefinitely in space and time, a universe vastly greater than the observable region, seems to imply that red-shifts are not primarily velocity-shifts.

In view of this possible conflict, whether of facts or theories or speculations, the observer is inclined to keep an open mind and to adopt parallel working hypotheses for the interpretation of his explorations. He may assume, first, that red-shifts are velocity-shifts, or, secondly, that red-shifts result from some unknown principle that does not involve actual motion, and



The arrows above the nebular spectra point to the H and K lines of calcium and show the amounts these lines are displaced toward the red end of the spectra. The comparison spectra are of helium.

The direct photographs (on the same scale and with approximately the same exposure times) illustrate the decrease in size and brightness with increasing velocity or red-shift.

NGC 4473 is a member of the Virgo cluster and NGC 379 is a member of a group of nebulae in Pisces.

always, of course, he will search for some empirical, critical test for distinguishing between the two assumptions, between motion and no motion.

Now let us examine the two assumptions. The true nature of light is still uncertain. For many purposes it can be discussed in two quite different ways, both of which lead to the same conclusions. We may picture light as spreading out in waves, or as travelling in bullet-like parcels of energy called quanta. Waves have various wave-lengths; quanta carry different amounts of energy. Since the same phenomenon can be logically described by both concepts, there is evidently some underlying connexion between them. This connexion is expressed in the very fundamental relation between the energy-content of a quantum, and the equivalent wave-length,

$$\text{energy} \times \text{wave-length} = \text{constant}$$
$$\therefore E\lambda = C.$$

Because of this relation an increase in λ implies a decrease in E, or, we may say, a reduction of E implies an increase in λ. In either case we observe only the increased wave-length, and we have no direct way of determining which of the two effects is fundamental. If the primary change is in the wave-length, then red-shifts are probably velocity-shifts. But the primary change might possibly be a loss of energy, which we would observe as a red-shift. In the latter case the law of red-shifts would be fully described by the simple statement that light loses energy in proportion to the distance it travels through space. Thus, on general considerations, we recognize alternative possibilities. Let us examine them more closely, beginning with velocity-shifts.

Red-Shifts as Velocity-Shifts

Relative motion in the line of sight shifts the lines in the spectrum of any light-source. If a luminous

body is rushing towards the observer, the light-waves are crowded on to one another, and are all shortened. Consequently, the entire pattern of spectral lines is displaced from its normal position, in the direction of the violet end of the spectrum—towards the shorter waves. On the other hand, if the light-source is rapidly receding from the observer, the light-waves are dragged out and lengthened, and the pattern of the spectrum is shifted towards the red.

In both cases, whether the shift $d\lambda$ is to the red or to the violet, the fractional displacement $d\lambda/\lambda$ is constant throughout a given spectrum. Each line is shifted by a constant fraction of its normal wave-length. Thus the fraction describes and measures the motion in the line of sight, the 'radial velocity' as it is called. Actually, the velocity of the light-source v is merely the velocity of light c multiplied by the fractional velocity-shift,[1]

$$v/c = d\lambda/\lambda$$
$$v = c \times d\lambda/\lambda.$$

For instance, a shift of 0·00001, or one part in a hundred thousand, represents a velocity of 1·86 miles per second; a shift of 0·001, or one-tenth of 1 per cent., a velocity of 186 miles per second.

Velocity-shifts of these dimensions, both to the violet and to the red (approaching and receding), are well known in the laboratory, and among the planets and the stars; their study, in fact, is an essential part of the investigation of stellar motions. And occasionally, in the tremendous explosions we observe as novae, masses of gas are driven away from a star at even greater velocities; violet-shifts have been recorded as large as 1 per cent., representing velocities of approach

[1] The relation $v/c = d\lambda/\lambda$ is a first approximation which serves well enough for small shifts. The rigorous expression, derived from the theory of relativity, is complicated and departs more and more from the simple relation as the shifts increase.

as high as 1,860 miles per second. Thus velocity-shifts, on a microscopic scale, are familiar phenomena, and their interpretation is not to be questioned.

Now the red-shifts observed in nebular spectra behave as velocity-shifts behave—the fractional shift $d\lambda/\lambda$ is constant throughout a given spectrum—and they are readily expressed as velocities of recession. The scale is so convenient that it is widely used, even by those cautious observers who prefer to speak of 'apparent velocities' rather than actual motion. For instance, the law of red-shifts is frequently called the 'velocity-distance relation'.

When Slipher, in his great pioneering work, assembled the first considerable lists of red-shifts, the observations were necessarily restricted to the brighter, nearer nebulae. Consequently, the shifts were moderately small (less than 1 per cent.), and they were accepted without question as the familiar velocity-shifts. Attempts were immediately made to study the motions of the nebulae by the same methods used in the study of stellar motions. But later, after the 'velocity-distance relation' had been formulated, and Humason's observations of faint nebulae began to accumulate, the earlier, complete certainty of the interpretation began to fade.

The disturbing features were the facts that the 'velocities' reached enormous values and were precisely correlated with distance. Each million light-years of distance added a hundred miles per second to the 'velocity'. As Humason swept farther and farther out into space he reported 'velocities' of 5,000 miles per second, then 10,000 then 15,000. Finally, near the absolute limit of his spectrograph he recorded red-shifts of 13 and 14 per cent., 'velocities' of about 25,000 miles per second—around the earth in a second, out to the moon in 10 seconds, out to the sun in just over an hour.

Red-shifts continue to increase beyond the range of the spectrograph, and, for the faintest nebulae that can be photographed, they are presumably about double the largest recorded shifts—the 'velocities' are about 50,000 miles per second. These quantities we are asked to accept as measuring a general recession of the nebulae, an expansion of the universe itself. The law of red-shifts then reads: the nebulae are receding from the earth, in all directions, with velocities that are proportional to their distances from the earth.

Red-Shifts as Loss of Energy in Transit

Well, perhaps the nebulae are all receding in this peculiar manner. But the notion is rather startling. The cautious observer naturally examines other possibilities before accepting the proposition even as a working hypothesis. He recalls the alternative formulation of the law of red-shifts—light loses energy in proportion to the distance it travels through space. The law, in this form, sounds quite plausible. Internebular space, we believe, cannot be entirely empty. There must be a gravitational field through which the light-quanta travel for many millions of years before they reach the observer, and there may be some interaction between the quanta and the surrounding medium. The problem invites speculation, and, indeed, has been carefully examined. But no satisfactory, detailed solution has been found. The known reactions have been examined, one after the other—and they have failed to account for the observations. Light *may* lose energy during its journey through space, but if so, we do not yet know how the loss can be explained.

The observer seems to face a dilemma. The familiar interpretation of red-shifts leads to rather startling conclusions. These conclusions can be avoided by an assumption which sounds plausible but which finds no

place in our present body of knowledge. The situation can be described as follows. Red-shifts are produced either in the nebulae, where the light originates, or in the intervening space through which the light travels. If the source is in the nebulae, then red-shifts are probably velocity-shifts and the nebulae are receding. If the source lies in the intervening space, the explanation of red-shifts is unknown but the nebulae are sensibly stationary.

The Critical Test between Alternative Interpretations

The distinguishing feature between the two pictures is the recession. So the observer concentrates on the recession. He inquires if there is an empirical test for determining whether or not a luminous body is rapidly receding. The answer is *yes*; in principle, at least, a test does exist. A rapidly receding nebula should appear fainter than a stationary nebula at the same momentary distance. If we had previous knowledge of the distances of nebulae, an examination of their apparent faintness would tell us at once whether or not they are receding.

The test is valid in principle but, unfortunately, there is a catch in its application to the particular problem. Our only information concerning great distances is derived from the same apparent luminosities which we desire to test. If a nebula appears abnormally faint because of recession, the fading would merely introduce an error in our estimation of distance, and the two effects could not be distinguished. Perhaps later, if apparent diameters can be calibrated as reliable measures of distance, or new criteria can be formulated, it may be possible to apply the test directly. At present, however, the direct investigation ends in a vicious circle, and the persistent observer is forced to consider a possible indirect attack on the problem.

The Indirect Test

The point of departure for the indirect attack is the scale of distance as determined from apparent luminosities. There are evidently two possible scales, depending upon whether or not the nebulae are receding. The distinction is not merely a difference in the unit; the ratios of the distances on the two scales do not remain constant. Consequently, the wrong scale may lead to peculiarities or inconsistencies in the map of the observable region, which might be recognized, or, at least, suspected. The project of mapping the observable region in detail cannot be carried out, but it is possible to examine precise formulations of the general characteristics of the region. Actually, the investigation reduces to a study of the laws of red-shifts and the laws of nebular distribution as derived from the alternate scales of distance.

Effects of Red-Shifts on Apparent Luminosity

The first step in the venture is the calculation of the proper corrections which must be applied to the measured apparent luminosities before the distances can be estimated. There are two possible corrections, the reasons for which are as follows. The apparent luminosity of a nebula is measured by the rate at which the radiant energy reaches the observer. The observer intercepts light-quanta, each carrying a certain amount of energy. A change in either the rate of arrival of the quanta, or in the individual energies they carry, will alter the measured luminosity. These considerations introduce two possible corrections, either or both of which must be applied according to circumstances.

The rate of arrival, the number of quanta which the observer intercepts each second, is necessarily reduced by recession of the nebulae. The appropriate correc-

tion, known as the *recession factor*, must be applied if, and only if, red-shifts are velocity-shifts. Here, in principle, lies the empirical test of the interpretation of red-shifts, which will be discussed later in considerable detail.

The Energy Effect

Meanwhile it may be pointed out that the individual energies of the quanta reaching us from the nebulae are altered by red-shifts, quite apart from the question of recession. In view of the fundamental relation previously mentioned,

$$\text{energy} \times \text{wave-length} = \text{constant},$$

the mere presence of red-shifts, the observed increase in wave-lengths, necessarily implies a reduction in the energies. This *energy effect* operates regardless of the interpretation of red-shifts. It must be applied as a matter of course to all measured luminosities before distances can be estimated.

The calculation of the energy effect starts from the fact that the fractional red-shift $d\lambda/\lambda$ is constant throughout a given spectrum. Each wave-length is increased, and, consequently, the energy in each quantum is reduced, by the constant factor $1+d\lambda/\lambda$. We might naturally conclude that the total luminosity, the sum of all the individual energies, would be reduced by the same factor. The conclusion, in fact, is valid in the particular case of the 'bolometric' luminosity, which is the total radiation, integrated over all wave-lengths, as it would be measured in empty space outside the earth's atmosphere. But we are concerned with the photographic luminosity. Hence the effect must be traced through the atmosphere, the telescope, and the photographic film. The procedure is complicated because the energy effect is selective (varying through the spectrum). The distortions magnify the original factor

$1+d\lambda/\lambda$ to a factor that approaches $1+3d\lambda/\lambda$. More precisely, the photographic magnitudes, to use the technical measure of apparent faintness, are increased by the increment,

$$\Delta m = 3d\lambda/\lambda.$$

This is the correction for the energy effect which is valid regardless of the origin of red-shifts,[1] and which is applied as a matter of course to all measures of apparent luminosity used for the estimation of distances.

The Recession Factor

The *recession factor* reduces the rate at which the quanta reach the observer because it dilutes the stream of quanta. The effect is quite simple, and may be described in the following way. Suppose that, in one second of time, a stationary nebula radiates a certain number of quanta n in the direction of the observer. Since all of the quanta travel with the velocity of light, c miles per second, the first of the quanta to leave the nebula had travelled forward the distance $c = 186{,}000$ miles, when the last of the n quanta leaves the nebula. Thus the n quanta which leave the nebula during one second are scattered uniformly over a path whose length is c.

Now suppose that the nebula is receding from the observer with a velocity of v miles per second. It still radiates the same number of quanta in one second of time, namely n, but during the interval between the emission of the first and the last of these quanta the

[1] Since $E\lambda$ = constant = h/c, the above statement assumes that the relation between Planck's constant h and the velocity of light c is essentially the same for light from the distant nebulae as in light from laboratory sources. The assumption has been tested by measuring red-shifts both with a grating spectrograph and a prism spectrograph. The fact that the shifts are the same in both cases seems to confirm the constancy of h/c. (*Annual Report of the Mount Wilson Observatory*, 1935–6.)

first quantum has travelled forward the distance c while the nebula has receded the distance v. Thus the n quanta from the receding nebula are distributed over the path $c+v$, while the same number of quanta from the stationary nebula are distributed over the path c. Evidently, the recession has increased the separation between successive quanta by the factor $(c+v)/c = 1+v/c$. The dilution of the stream of quanta necessarily reduces their rate of arrival and, consequently, the apparent luminosity of the nebula. The recession factor $1+v/c$ must be applied to a receding nebula, in addition to the energy factor; the observer receives fewer quanta and each of the quanta has lost energy.

In practice, the expression $1+v/c$ is generally replaced by $1+d\lambda/\lambda$, because the use of the factor implies an expanding universe, and current theories of such a universe regard the second expression as the proper formulation. This procedure is followed in the present discussion.

The recession effect, unlike the energy effect, is not selective; it is not distorted by the atmosphere, the telescope, or the photographic film. Consequently, the recession factor, in its simple form $1+d\lambda/\lambda$, may be applied to apparent luminosities on any system, photographic as well as bolometric. The fractional reduction in luminosity is merely the fractional red-shift. In technical terms apparent magnitudes are increased by the increment

$$\Delta m = d\lambda/\lambda.$$

The effect of recession is quite negligible for ordinary velocities (or red-shifts); for instance, a velocity of 1,860 miles per second (red-shift = 0·01) reduces the apparent luminosity by only 1 per cent. But at great distances, where red-shifts are large, the effects are measurable quantities. For the largest shifts recorded

with the spectrograph the reduction is about 14 per cent.; at the limit of the deepest accurate survey, the estimated reduction is about 23 per cent.

It is evident that critical tests of the interpretation of red-shifts as velocity-shifts could be readily made, provided we had reliable distance-criteria which were independent of apparent luminosity. But in the absence of such criteria we must return to the method of parallel hypotheses, and examine the general features of the observable region on the basis of alternative scales of distance.

The Alternative Laws of Red-Shifts

(a) *Red-Shifts not interpreted as Velocity-Shifts*

Let us start with the precise formulation of the law of red-shifts. The best data for the purpose are those found in the great clusters of nebulae, for nine or ten of which reliable red-shifts and distances have been derived from the brighter members. The distances range from about 7 million to 240 million light-years, the latter being the greatest reliable distance that has been assigned to an individual object. If peculiarities in the behaviour of red-shifts can be detected at the present time, they will be found in this group of data.

The observational data are quite simple; they consist merely of measured red-shifts and of measured luminosities, corrected for energy-effects. The luminosities indicate the distances, which, of course, are expressed on two scales according to whether or not the recession factors are introduced. Except for the choice of scales, the relative distances are reliable.

Now let us adopt the first alternative and assume that red-shifts are not velocity-shifts, that the universe is not rapidly expanding. In this case the red-shifts are strictly proportional to the distances, over the entire range of the data. Within the very small uncer-

tainties of the measures the law of red-shifts is strictly linear as far out into space as spectra have been recorded. The law is expressed as follows:

$$\text{red-shifts} = \text{constant} \times \text{distance}$$
$$d\lambda/\lambda = kr.$$

The implications of this result are important. If red-shifts are not velocity-shifts, light loses energy strictly in proportion to the distance it travels through

TABLE I

Observational Data indicating the Law of Red-Shifts

Cluster	$d\lambda/\lambda$	log $d\lambda/\lambda$	m_c*
Virgo	0·0041	−2·387	10·49
Pegasus	0·0127	1·900	12·88
Perseus	0·0174	1·759	13·48
Coma	0·0245	1·611	14·23
U. Ma. I†	0·0517	1·286	16·12†
Leo	0·0653	1·185	16·33
Cor. Bor.	0·0707	1·151	16·54
Boötes	0·1307	0·884	17·89
U. Ma. II†	0·1403	−0·853	17·73†

* m_c is the apparent magnitude of the fifth brightest nebula in the cluster, corrected for local obscuration and for energy-effects of red-shifts. The distance of a cluster is indicated by the relation, log $r = 0{\cdot}2m_c + 4{\cdot}29$.

† The data for the two Ursa Major clusters are the most uncertain of those in the table, because the red-shifts are each derived from a single spectrogram of a single nebula.

The data are taken from *Contributions of the Mount Wilson Observatory*, No. 549; *Astrophysical Journal*, **84**, 270, 1936.

space. As light streams in from the remote nebulae in all directions, each million years of the light-paths subtracts the same fraction of energy from the quanta. We may not know how the reduction is accomplished, but we do know that the action is everywhere uniform. Within the small uncertainties of the measures the sample of the universe that can be explored with spectrographs is thoroughly homogeneous.

There are no systematic departures from uniformity whose trends can be pushed out into the universe at large. The homogeneity may extend indefinitely, or trends may exist which only become appreciable beyond

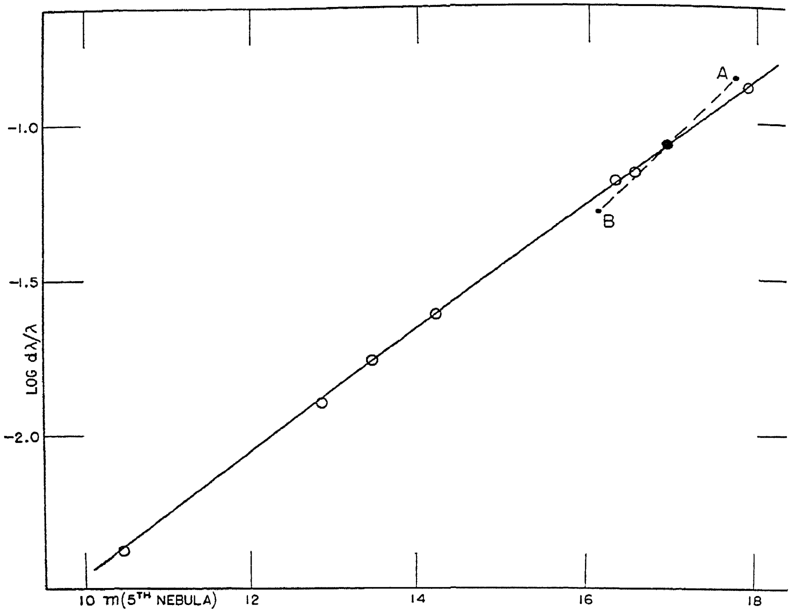


FIG. 1. The Linear Law of Red-Shifts not interpreted as Velocity-Shifts.

Each point represents the mean red-shift for a great cluster together with the distance indicator, namely, the apparent magnitude of the fifth brightest nebula, corrected for energy effects of the red-shifts. The data are found in Table I. The points *A* and *B*, referring to the two Ursa Major clusters, have small weight, and are combined into the mean point represented by the large black disk.

The points define the straight line,

$$\log d\lambda/\lambda = 0{\cdot}2m - 4{\cdot}457,$$

whose slope establishes the linearity of the law of red-shifts out as far as spectra have been recorded.

The inclusion of recession factors would displace all the points to the left by amounts which increase from left to right (with increasing m), thus destroying the linearity of the law of red-shifts. The displacements are not conspicuous on the scale of the diagram, and they are shown separately in Fig. 2.

the range of the spectrograph; on these questions our knowledge fails and we are reduced to mere speculation. If the explored region is a fair sample, it is relatively insignificant, and it furnishes only negative information concerning a universe that must be vastly greater. This conclusion seems plausible and, in a sense, familiar.

(b) *Red-Shifts interpreted as Velocity-Shifts*

But suppose that red-shifts are velocity-shifts, and that the nebulae are receding. Then the law of red-shifts furnishes information, not concerning the contents or properties of internebular space, but concerning the expansion of the universe as a whole. There seems to be no *a priori* necessity for a linear law of expansion, a strict proportionality between red-shifts and distance. Indeed, the general theory indicates that the law must depart from linearity. If our sample is sufficiently large we may, perhaps, observe the departures and determine their trend. The information would immediately restrict our actual world to a particular family of possible worlds.

With this end in view let us re-examine the data for the great clusters, applying to the luminosities the additional correction for recession. The corresponding distances, of course, are systematically altered but the distortions are not uniform. Distances as previously estimated, without the recession factor, are now shortened, and by percentages which increase with the distance. For instance, the distance to the nearest cluster is reduced by less than one-fifth of 1 per cent., while the distance of the most remote cluster is reduced by about 7 per cent. The red-shifts, however, remain unchanged.

Consequently, the law of red-shifts, which was linear when the recession factor was omitted, cannot be linear

when the recession factor is included. We now observe a small but systematic trend away from the linear law. These departures are our first positive clue to the structure of the universe. If the universe is expanding, we know something about how it expands. Our sample is large enough to be significant.

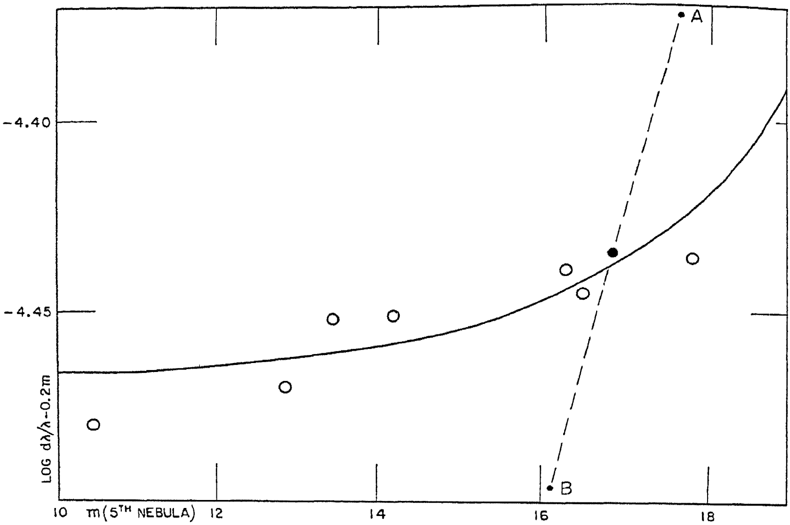


FIG. 2. Departures from Linearity in the Law of Red-Shifts interpreted as Velocity-Shifts.

The points refer to the same clusters as those in Fig. 1. The red-shifts are unchanged but the magnitudes have been further corrected by the recession factors. The vertical scale represents departures from a linear law of red-shifts. For a linear law, $\log d\lambda/\lambda = 0{\cdot}2m + \text{constant}$, or

$$\log d\lambda/\lambda - 0{\cdot}2m = \text{constant},$$

and the points in the diagram would cluster around a horizontal line. Actually, the quantities $\log d\lambda/\lambda - 0{\cdot}2m$ vary systematically with m, or, in other words, with distance. The precise form of the departure-distance relation is not indicated by the points—for instance, they would fit a linear relation well enough—and the curve drawn in the diagram is that corresponding to the power series

$$d\lambda/\lambda = kr + lr^2 + \ldots,$$

by which the law of red-shifts is commonly expressed in relativistic cosmology.

As in Fig. 1 the two points A and B are combined into the single mean point represented by the large black disk.

The departures are positive—in other words, the red-shift for a given distance is larger than the linear law would lead us to expect. Moreover, the departures, the differences between the observed law and the linear law, increase with distance at an accelerated rate. The new law may be written by adding a term to the old law; thus

$$d\lambda/\lambda = kr + \text{departures},$$

where the departures are some function of the distance r. The actual observations, because of the small, residual uncertainties of the measures, might be represented well enough by many different functions of the distance. However, we will follow the custom of relativistic cosmology, and use a power series as a first approximation to the real function. Then the series

$$d\lambda/\lambda = kr + lr^2 + mr^3 + \ldots$$

represents the non-linear law of red-shifts that follows from the assumption that red-shifts are velocity-shifts.

The first term kr is the linear relation previously derived without the correction for recession. Expressed in terms of velocities the value of k represents a rate of increase of about 100 miles per second per million light-years of distance. Before the new investigations were made, the 'velocity-distance relation', as the law of red-shifts was often called, included this first term alone. The relation was known to be approximately linear and the precision of the data did not justify the addition of the small, higher order terms. The linear term, when followed out into space, suggested that the velocity of light itself would be reached at a distance of about 1,860 million light-years, or less than four times the average distance of the faintest nebulae recorded with the 100-inch reflector.

Moreover, the receding nebulae could be traced backward in time to a very remarkable epoch, about

1,860 million years ago. At that epoch the nebulae would all have been found in our immediate neighbourhood. The present distribution would be reproduced if all the nebulae, being then in a compact cluster, had suddenly started to recede at different velocities. The fastest would now be the most distant, while the slowest would still be in our vicinity. Evidently, the present distances would be proportional to the velocities. The initial instant, the famous t_0, is back in time about 1,860 million years. In the favoured, expanding worlds of relativistic cosmology, the period was generally called 'the age of the universe', the time-interval since the expansion began.

Efforts were made to crowd a vast multitude of events into the brief span, but serious difficulties were encountered. These difficulties often led to the expectation that, when departures from the approximate linear 'velocity-distance relation' were finally detected, they would prove to be negative. In other words, it was expected, or hoped, that further investigations would show that the expansion began very slowly but continually accelerated until it reached the enormous speed we observe to-day. Consequently, the true age of the universe would be much greater than that estimated from the provisional linear relation.

Well, we have the first intimation of the nature of the departures—and they are positive. Long ago when the light we now record left the very distant nebulae, the expansion was more rapid than it is to-day. The time-scale is not lengthened; on the contrary, it is materially shortened. This feature of the expansion is contained in the second term of the power series representing the non-linear law of red-shifts, namely, the term lr^2. Actually, the third, and higher, order terms should be included with the second term, but they are presumably so small that they can be neglected in the

present discussion. Within the range of existing spectrographs, at any rate, they are smaller than the uncertainties of the measures.

The coefficient l is definitely a positive quantity, and its value is about $2{\cdot}5 \times 10^{-19}$ (light-years)$^{-2}$. Expressed in terms of velocities it corresponds to about one-twentieth of a mile per second at a distance of a million light-years. The term seems insignificant, and certainly it does not become appreciable until great distances are reached. Thereafter, it increases rapidly, for the coefficient is multiplied by the square of the distance. At 100 million light-years the term adds about 500 miles per second to the linear term, and at 500 million light-years about 12,500 miles per second.

The chief significance of the term for cosmological theory lies in the positive sign. The rate of expansion of the universe has been slowing down, at least for the past several hundred million years. The 'age of the universe' is considerably shorter than that permitted by the linear law, although the precise, numerical reduction cannot be determined until the third and higher terms of the power series have been evaluated. The maximum permissible span appears to be of the order of 1,500 million years, but the true value might lie anywhere between the maximum and half the maximum. The initial instant, the t_0, clearly falls within the life-history of the earth, probably within the history of life on the earth. And, as we look back into time with our telescopes, we pass in review from a half to a third of the entire period of expansion.

The Dilemma

Thus the familiar interpretation of red-shifts as velocity-shifts leads to strange and dubious conclusions; while the unknown, alternative interpretation leads to conclusions that seem plausible and even

familiar. Theories may be revised, new information may alter the complexion of things, but meanwhile we face a rather serious dilemma. Some there are who stoutly maintain that the earth may well be older than the expansion of the universe. Others suggest that in those crowded, jostling yesterdays, the rhythm of events was faster than the rhythm of the spacious universe to-day; evolution then proceeded apace, and, into the faint surviving traces, we now misread the evidence of a great antiquity. Our knowledge is too meagre to estimate the value of such speculations, but they sound like special pleading, liked forced solutions of the difficulty. The fundamental question is the interpretation of red-shifts.

REMOTE CLUSTERS

The two clusters reproduced in the plate are extremely remote and, consequently, rather inconspicuous. Their interest lies in the fact that they bracket the present range of the spectrograph.

a. The Boötes cluster, at 240 million light-years, has furnished the largest, reliably measured red-shift, $d\lambda/\lambda = 0{\cdot}1307$, which corresponds with a velocity of recession of about 24,200 miles per second. The area reproduced in the plate is about 56 square minutes of arc (about 8 per cent. of the area of the full moon), and contains about one hundred nebulae, two-thirds of which are members of the cluster. The large nebula, N.G.C. 5672, is relatively near our own system but happens to lie in the direction of the cluster.

b. The Hydra cluster, 5′ north and 2·7′ east of N.G.C. 2716, appears considerably fainter than the Boötes cluster because, although it is only slightly more remote, it lies in lower galactic latitude (+30°) and is more affected by local obscuration. Attempts to measure the red-shift in the Hydra cluster have hitherto failed. The reproduction, greatly enlarged from the original negative, shows about 70 nebulae in an area of 22 square minutes of arc.

CHAPTER III

POSSIBLE WORLDS

THE previous lecture described the appearance and behaviour of red-shifts in the spectra of nebulae, and called attention to the alternative possible interpretations. If red-shifts are produced in the nebulae, where the light originates, they are probably the familiar velocity-shifts, and they measure an expansion of the universe. If the nebulae are not rapidly receding, red-shifts are probably introduced between the nebulae and the observer; they represent some unknown reaction between the light and the medium through which it travels.

In principle, at least, it is possible to distinguish between the alternatives, because a rapidly receding nebula should appear fainter than a stationary nebula at the same distance. The test cannot be applied directly, for apparent faintness, which we wish to compare with distance, is itself our only measure of great distance. However, the problem may be approached indirectly. The parallel assumptions of motion and no motion lead to different distances for nebulae of the same measured apparent luminosity. Thus there are alternative scales of distance, and they furnish quite different conceptions of our sample of the universe. It is possible that the wrong scale may introduce anomalies which can be detected or at least suspected.

With this end in view, the effects of recession on apparent luminosity, and, consequently, on estimated distances, were calculated, and the laws of red-shifts were formulated precisely both with and without the recession factors. The assumption of no motion led to a strictly linear law which indicates that the region of space explored with spectrographs is thoroughly

homogeneous. This result implies that the explored region is too small a sample to furnish positive information concerning the universe at large. We merely infer that the universe is vastly greater than the sample, and the conclusion seems quite plausible.

The assumption of motion, on the other hand, led to a non-linear law of red-shifts, according to which the velocities of recession accelerate with distance or with time counted backward into the past. A universe that has been expanding in this manner would be so extraordinarily young, the time-interval since the expansion began would be so brief, that suspicions are at once aroused concerning either the interpretation of red-shifts as velocity-shifts or the cosmological theory in its present form.

Surveys of Nebulae

The present lecture describes similar results which are derived from the laws of nebular distribution as formulated with the alternative scales of distance. The preliminary reconnaissance had indicated that the large-scale distribution is at least roughly uniform. Guided by this information accurate surveys were planned for the purpose of formulating precisely the law of distribution. Five such surveys have now been reported, made with three large reflectors, by two observers working independently. One survey was made by Dr. Mayall with the 36-inch reflector, at the Lick Observatory, the others were made at the Mount Wilson Observatory, two with the 60-inch reflector and two with the 100-inch. The results are internally consistent, and the diversity of the investigations offers some assurance against the presence of hidden systematic errors—that nightmare of the cautious observer—which might vitiate the conclusions.

Each survey, although it required many months or

some years for completion, is summarized by a single symbol N_m, which represents the average number of nebulae per unit area, brighter than a specified limit of apparent faintness. It is unnecessary to describe the simple but laborious methods by which the nebular counts were reduced to a standard, homogeneous system, and the limits of apparent faintness determined with considerable accuracy. It is sufficient to say that, in their final forms, the surveys represent about 100,000 nebulae, and that the distances corresponding to the limits of the surveys range from about 150 to 400 million light-years. The surveys, it will be noticed, extended far back into time as well as far out into space.

Analysis of the surveys consists in the comparison of numbers of nebulae with the volumes of space they occupy. If the large-scale distribution is uniform, the numbers will evidently be proportional to the volumes. If the proportionality does not emerge from the data, we must conclude that the observable region is not homogeneous. The discrepancies must then be examined in order to find whether they represent random variations within the sample, or departures from uniformity which vary systematically with distance. Systematic departures would be important because the trend could be pushed out beyond the limits of the telescope to furnish positive information concerning the universe at large. On the other hand, homogeneity, or random variations, would suggest that our sample is a small, insignificant portion of the universe.

Now let us follow this programme of investigation. Each survey indicates the number of nebulae in the sky that are brighter than a certain limit of apparent faintness (corrected for local obscuration within the stellar system). Apparent faintness can be transformed into distances on either of two possible scales. Consequently, each survey furnishes a certain number of

nebulae distributed through either of two spheres with specified radii and specified volumes.

When the surveys are compared, the crude observations, uncorrected for red-shifts—the World Picture, to use Professor Milne's happy phrase—indicate that the number of nebulae increase less rapidly than the volumes of space through which they are distributed; the nebulae *appear* to thin out with distance. However, the measured luminosities must be corrected for energy-effects, regardless of the interpretation of red-shifts. The corrections will reduce the estimated distances and volumes, and, consequently, will tend to compensate the apparent thinning out.

TABLE II

Surveys of Nebulae

m	$d\lambda/\lambda$	$\log N$
20·34	0·231	3·162
19·53	0·158	2·686
19·03	0·125	2·342
18·68*	0·106	2·161
18·11	0·086	1·886

m is limiting magnitude actually observed, corrected for local obscuration and for energy-effects; $d\lambda/\lambda$ is the red-shift corresponding to the limiting magnitude m; $\log N$ is the logarithm of the average number of nebulae per square degree brighter than magnitude m.

* The fourth survey was made by N. U. Mayall using the 36-inch reflector at the Lick Observatory.

The five surveys are discussed in *Contributions of the Mount Wilson Observatory*, No. 557; *Astrophysical Journal*, 84, 517, Dec. 1936.

The energy-corrections lead to the first of the two scales of distance, namely, that which follows from the assumption that red-shifts are not velocity-shifts. On this scale the law of red-shifts is strictly linear over an observed range of about 250 million light-years, and, presumably, will continue to be approximately linear over a considerably greater range. Therefore, the red-shifts at the limits of the various surveys are readily

ascertained, and uncertainties are restricted to the deepest survey alone. The energy-corrections, as previously explained, are simple functions of the red-shifts.

The Law of Nebular Distribution when Red-Shifts are not interpreted as Velocity-Shifts

We can now formulate the law of nebular distribution on the assumption that the nebulae are not rapidly receding.[1] The energy-corrections corresponding to red-shifts at the limits of the various surveys exactly compensate for the apparent thinning out in the World Picture. Numbers of nebulae are strictly proportional to the volumes of space they occupy. If the universe is not rapidly expanding, the observable region is thoroughly homogeneous out as far as accurate surveys have been extended.

This important conclusion is derived by comparing simple counts of nebulae with measured luminosities, corrected by energy-factors which must be applied to such measures regardless of cosmological theory. The uniform distribution is a plausible and welcome result. Homogeneity within our sample seems so plausible, in fact, that it has often been adopted as a preliminary assumption. The *apparent* distribution indicated by the crude observations is then described as a true, uniform distribution plus *apparent* departures from uniformity. The apparent departures can be expressed as corrections to the measured luminosities of nebulae

[1] The linear law of red-shifts (omitting recession factors) is $d\lambda/\lambda = kr$, and may be written

$$\log_{10} d\lambda/\lambda = 0{\cdot}2(m_0 - 3d\lambda/\lambda) + \text{constant},$$

where $d\lambda/\lambda$ is the mean red-shift for nebulae of apparent magnitude m_0; m_0 is the measured magnitude corrected for local obscuration; the constant, which involves k and the intrinsic luminosity of nebulae (candle-power), is $-4{\cdot}707$. From this equation red-shifts and the corresponding energy-effects $\Delta m = 3d\lambda/\lambda$ are readily derived for the limiting magnitude of each survey.

at the limits of the various surveys, and these corrections, in turn, can be compared with the red-shifts at the same limits. The result, expressed in terms of magnitude increments Δm, is

$$\Delta m = 2{\cdot}94 d\lambda/\lambda,$$

which agrees very closely with the energy-corrections

$$\Delta m = 3 d\lambda/\lambda.$$

The energy-corrections, it will be recalled, are the *total* effects of red-shifts on apparent luminosities, provided red-shifts are not velocity-shifts. The latter interpretation seems to follow directly from the preliminary assumption of uniformity.

The assumption of uniformity has much to be said in its favour. If the distribution were not uniform, it would either increase with distance, or decrease. But we would not expect to find a distribution in which the density increases with distance, symmetrically in all directions. Such a condition would imply that we occupy a unique position in the universe, analogous, in a sense, to the ancient conception of a central earth. The hypothesis cannot be disproved but it is unwelcome and would be accepted only as a last resort in order to save the phenomena. Therefore, we disregard this possibility and consider the alternative, namely, a distribution which thins out with distance.

A thinning out would be readily explained in either of two ways. The first is space absorption. If the nebulae were seen through a tenuous haze, they would fade away faster than could be accounted for by distance and red-shifts alone, and the distribution, even if it were uniform, would appear to thin out. The second explanation is a super-system of nebulae, isolated in a larger world, with our own nebula somewhere near the centre. In this case the real distribution would thin out after all the proper corrections had been applied.

Both explanations seem plausible but neither is permitted by the observations. The apparent departures from uniformity in the World Picture are fully compensated by the minimum possible corrections for red-shifts on any interpretation. No margin is left for a thinning out. The true distribution must either be uniform or increase outward, leaving the observer in a unique position. But the unwelcome supposition of a favoured location must be avoided at all costs. Therefore, we accept the uniform distribution, and assume that space is sensibly transparent. Then the data from the surveys are simply and fully accounted for by the energy corrections alone—without the additional postulate of an expanding universe.

In this case all the empirical information we have concerning the observable region as a whole is internally consistent. The region appears to be thoroughly homogeneous—an insignificant sample of a universe which extends indefinitely. The conclusion would probably be accepted without hesitation if it were not for the fact that, at the moment, we do not know of any permissible interpretation of red-shifts other than actual motion, actual recession of the nebulae.

The Problem of Distribution in an Expanding Universe

Therefore, we must consider the alternative scale of distance, and formulate the law of distribution on the assumption that red-shifts are the familiar velocity-shifts, and do measure the expansion of the universe. The actual recession necessitates one correction to apparent luminosities and another to the epoch of the various surveys. The first correction has already been discussed. Recession reduces apparent luminosities of the nebulae by the 'recession factors' $1+d\lambda/\lambda$. When these effects are removed from the measures, the nebulae appear brighter, and the distances are less than those

estimated from the uncorrected measures. Now the latter data, as we have just seen, indicate uniform distribution. Consequently, the revised distances might be expected to introduce departures from uniformity, in the sense that the volumes increase less rapidly than the numbers of nebulae, or, in other words, that the distribution increases outwards leaving the observer in an unwelcomed, favoured position. However, this conclusion does not necessarily follow, because the surveys represent different epochs in the history of the expanding universe.

The light which reaches us to-day left the limits of the various surveys far back in past time. From the limit of the deepest survey, for instance, the light started about 400 million years ago. It travelled for about 120 million years before it reached the limit of the next deepest survey, and another 130 million years before crossing the limit of the shallowest survey. During these immense intervals of time the nebulae at the limits of the different surveys were receding at enormous velocities to still greater distances.

We count a certain number of nebulae and we know that they were scattered through a certain volume of space *when the light left the limit of the survey*. But to-day, many millions of years later, these same nebulae are scattered through a much larger volume of space, and the increase is different for each survey. Evidently, all the surveys must be reduced to the same epoch before the law of distribution can be formulated. Moreover, the law will continually change, for the recession implies that the distribution thins out with time.

These considerations emphasize the complexity of the problem of distribution in an expanding universe. The first step in the solution is the choice of a common epoch to which the different surveys will be reduced in order to make the comparison of numbers of nebulae

and volumes of space. As a matter of convenience, the epoch selected is *now*, the time at which the surveys were made. Then, knowing the law of red-shifts or, in other words, the law of expansion, it seems possible to expand the volumes of all the surveys up to the epoch, now.

At this point the procedure becomes arbitrary. The calculations, in the present stage of knowledge, may be made in various ways, and the choice involves assumptions concerning the nature of the universe. As a simple illustration, does an individual nebula maintain a constant velocity as it recedes into the depths of space, or does its velocity steadily increase with increasing distance? This and other more technical questions must be answered before the reductions can be made with confidence. Thus the problem of reduction to a common epoch forces us to consider cosmological theory and some of the models of the universe which loom in that shadowy realm.

Most of the current models are derived from relativistic cosmology. Moreover, the outstanding exception, Professor Milne's kinematical model, is so outwardly similar, in several of its aspects, to a special case in the relativistic theory, that the observer, faced with a small sample, can scarcely hope to distinguish between them. Therefore, in the brief discussion which follows, the relativistic models alone will be considered. There are, it is said, many compelling reasons for concentrating on the theory, but the observer is not the proper authority to present them in their technical details. Instead, a few of the underlying principles will be mentioned together with the features of the models that may be compared with observations.

Expanding Universes of General Relativity

Relativistic cosmology is a natural offshoot of Einstein's theory of general relativity. However, the

cosmology is a superstructure, including other principles, and, if the present formulation were found to be inadequate, the failure would not necessarily affect the underlying theory. Relativity contributes the basic proposition that the geometry of space is determined by the contents of space. To this principle has been added another proposition, formulated in various ways and called by various names, but equivalent, in a sense, to the statement that all observers, regardless of their location, will see the same general picture of the universe. The second principle is a sheer assumption. It seems plausible and it appeals strongly to our sense of proportion. Nevertheless, it leads to a rather remarkable consequence, for it demands that, if we see the nebulae all receding from our position in space, then every other observer, no matter where he may be located, will see the nebulae all receding from his position. However, the assumption is adopted. There must be no favoured location in the universe, no centre, no boundary; all must see the universe alike. And, in order to ensure this situation, the cosmologist postulates spatial isotropy and spatial homogeneity, which is his way of stating that the universe must be pretty much alike everywhere and in all directions.

The kinds of universes that would be compatible with the relativity principle and the assumption of homogeneity have been determined by intricate mathematical reasoning. A body of necessary characteristics has been derived, one of which is of exceptional interest for our immediate problem, Such a universe, if it contains matter, will be unstable. At best it could be in unstable equilibrium, like a ball balanced on a point. The slightest disturbance would upset the balance—and internal disturbances evidently must occur. The universe would then revert to its natural state of either contraction or expansion. Theory does not indicate

either the direction or the rate of the change to be expected. The universe might be expanding or contracting and at a rate that is rapid or imperceptible. At this point the cosmologist seizes upon the observed red-shifts, interprets them as velocity-shifts, and presents them as visible evidence that the actual universe is now expanding, and expanding rapidly. It is for these reasons that relativistic cosmology is described as the theory of homogeneous, expanding universes which obey the relativistic laws of gravitation.

Another important aspect of such universes is the highly abstract concept of spatial curvature. The relativity principle states that the geometry of space is determined by the contents of space. Theoretical investigators, guided by the assumption of homogeneity, adopt Riemannian geometry which operates in curved space. The curvature cannot be visualized, and will not be discussed in detail. It is sufficient to say that the nature of the curvature is indicated, and the amount is measured, by the radius of curvature (which projects, as it were, into a higher dimension). The radius in our universe might be positive, negative, or zero, and might be large or small. A positive curvature implies closed space, a universe with a definite, finite volume but with no boundary. A negative curvature implies open space, an infinite universe. The limiting case of zero curvature is 'flat' Euclidean space with an infinite radius. Thus there are various types of curvature, and, in all but flat space, the amount of curvature has a wide range of possible values.

The general formulae which describe expanding universes include three arbitrary terms. These represent (*a*) the nature of the expansion, (*b*) the nature of the spatial curvature, and (*c*) the nature of the contents (to use a very loose conception of the so-called cosmological constant). Since the numerical values of the

terms may vary through wide ranges, the formulae present an infinite array of possible worlds. The problem of the observer, if the theory is valid, is to measure the three elements in our universe and thereby identify, among the possible models, the actual world we inhabit.

The first step has already been taken. For the law of red-shifts, the deviations from linearity introduced by the recession factors, has determined one of the three elements. If the universe is expanding, we now know how it expands. The law of nebular distribution, as we shall see, determines a second element, namely, the nature of the curvature.

Comparison of Observations with Theory

Now let us return to the surveys, and reduce them all to the epoch, *now*, in accordance with the principles of relativistic cosmology. We wish to know the relative numbers of nebulae which an observer, in an expanding universe, would count to successive limits of apparent faintness. The problem is intricate but it has been thoroughly investigated, and the necessary formula is available in quite simple terms.[1] Actually, the expression is just that previously derived for uniform distribution in a stationary universe, plus two extra terms. One of the terms represents the recession factor, the other represents effects of spatial curvature.

If the use of a logarithm is permitted, the situation may be clearly represented by a pair of equations. If nebulae are uniformly distributed through a non-expanding universe in which red-shifts are not primarily velocity-shifts, then the numbers should be propor-

[1] The requisite formulae have been derived by various investigators. Those used in the present discussion were developed by R. C. Tolman (*Relativity, Thermodynamics, and Cosmology*, Clarendon Press, (1934) and adapted to the specific problem of the surveys by Hubble and Tolman (Mt. Wilson Contr., No. 527; *Astrophysical Journal*, 82, 302, 1935).

tional to the volumes, and the surveys should conform (and actually do conform) with the relation

$$\log_{10} N = 0{\cdot}6m_c + \text{constant},$$

where m_c is the limiting faintness expressed as a magnitude, corrected for local obscuration and for the energy-effects required by the mere presence of red-

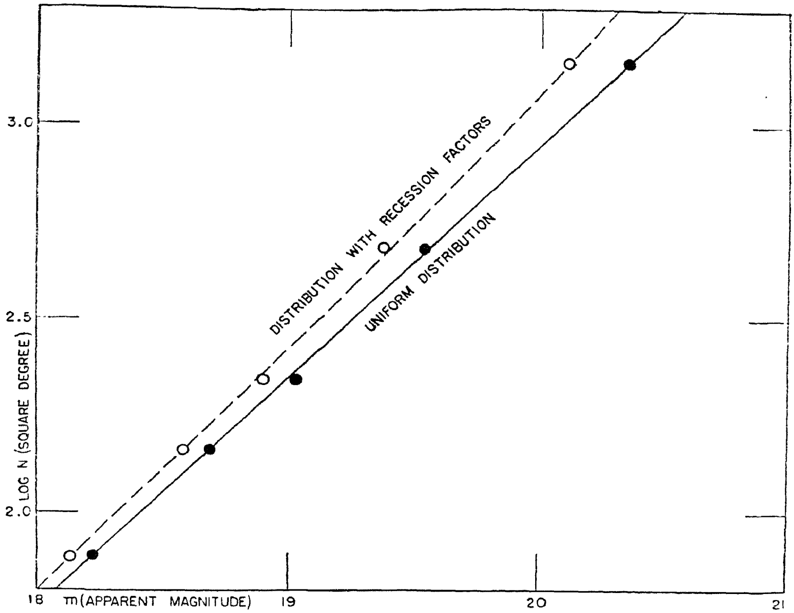


FIG. 3. The Law of Nebular Distribution on Alternative Interpretations of Red-Shifts.

Each point represents an entire survey and indicates the average number of nebulae per square degree, N, brighter than the limiting magnitude of the survey, m. The black disks represent the observed magnitudes corrected for local obscuration and energy-effects only. They define the straight line, $\log_{10} N = 0{\cdot}6m - 9{\cdot}05$, and indicate uniform distribution of nebulae in space when red-shifts are interpreted as not velocity-shifts.

The open circles represent the magnitudes further corrected by recession factors, and determine a distribution (the dash line) which departs systematically from uniformity. Homogeneity may be restored by postulating sufficient spatial curvature to offset the displacements introduced by the recession factors.

shifts.[1] The corresponding relation for a homogeneous, expanding universe, obeying the relativistic laws of gravitation, is

$$\log_{10} N = 0{\cdot}6(m_c - d\lambda/\lambda + C_v) + \text{constant},$$

where $d\lambda/\lambda$ is the recession factor and C_v is the effect of spatial curvature. We wish to know whether or not the surveys can be fitted into the latter expression.

If both of the extra terms (for recession and for curvature) were absent, the surveys would clearly fit the formula because the situation would be precisely that in a stationary universe. Now suppose we introduce only one of the extra terms, namely, the recession factor. In this way we pass from a stationary universe to an expanding universe with negligible curvature, but we destroy the agreement with the observations. The distribution is no longer uniform. The recession factors introduce departures from uniformity in the law of distribution, just as they introduced departures from linearity in the law of red-shifts.

Spatial Curvature

The departures from uniformity are positive; the numbers of nebulae increase faster than the volume of space through which they are scattered. Thus the density of the nebular distribution increases outwards, symmetrically in all directions, leaving the observer in a unique position. Such a favoured position, of course,

[1] The derivation is as follows. For uniform distribution numbers of nebulae N are proportional to volumes of space, and, consequently, to the cubes of the limiting distances r to which the counts are carried. Hence $\log_{10} N = 3 \log_{10} r + \text{constant}$. Among objects of the same candle-power the distances are proportional to the inverse square of the apparent luminosity l. Hence $\log_{10} r = \text{constant} - 0{\cdot}5 \log_{10} l$. Apparent magnitudes m measure apparent luminosities on a logarithmic scale. By definition, $m = \text{constant} - 2{\cdot}5 \log_{10} l$. Hence

$$\log_{10} r = 0{\cdot}2m + \text{constant}$$

and

$$\log_{10} N = 0{\cdot}6m + \text{constant}.$$

is intolerable; moreover, it represents a discrepancy with the theory, because the theory postulates homogeneity. Therefore, in order to restore homogeneity, and to escape the horror of a unique position, the departures from uniformity, which are introduced by the recession factors, must be compensated by the second term representing effects of spatial curvature.

There seems to be no other escape. Observations demonstrate that

$$\log_{10} N = 0{\cdot}6m_c + \text{constant}.$$

Relativistic cosmology requires that

$$\log_{10} N = 0{\cdot}6(m_c - d\lambda/\lambda + C_v) + \text{constant}.$$

Therefore, $$C_v = d\lambda/\lambda.$$

The curvature of space is demonstrated and measured by the postulated recession of the nebulae.

To the observer the procedure seems artificial. He has counted the nebulae to various limits, applied only the corrections that are necessarily required (energy-corrections), and derived the quite plausible result of uniform distribution. Now, in testing the relativistic theory, he introduces a new postulate, namely, recession of the nebulae, and it leads to discrepancies. Therefore, he adds still another postulate, namely, spatial curvature, in order to compensate the discrepancies introduced by the first. The accumulation of assumptions is uneconomical, and the justification must be sought in the general background of knowledge. The outstanding argument is the fact that velocity-shifts remain the only permissible interpretation of red-shifts that is known at the present time.

Well, perhaps the interpretation is correct and we do inhabit a rapidly expanding universe. In that case the surveys indicate the nature and amount of the spatial curvature. It must be such that the effects on the nebular distribution are just equal to the recession

factors at the limits of the various surveys. Actually, no curvature can be found which exactly compensates for the apparent departures from uniformity in each of the surveys. Nevertheless, if we admit the presence of rather considerable systematic errors in the observations, it is possible to select a curvature which will more or less restore the homogeneity. Hidden errors of the necessary dimensions are by no means impossible in the very delicate investigations near the limits of a great telescope.[1] They may be improbable, but they are not impossible. Therefore, the expanding universe can be saved by introducing a sufficient amount of spatial curvature. The plausible values are narrowly limited, and they indicate a radius of curvature that is positive and comparatively small. In fact the radius, about 470 million light-years, is a trifle less than the range of the 100-inch reflector for normal nebulae. Thus the second of the three arbitrary elements in the description of the relativistic universe is determined. If the universe is expanding, the spatial curvature is real and positive.

The nature of the curvature has rather grave implications. Since the curvature is positive, the universe is closed. Space is closed as the surface of a sphere is closed. The universe has a definite, finite volume although it has no boundaries in three-dimensional space. The remarkably small numerical value of the radius of curvature is a complete surprise. It implies that a large fraction of the universe, perhaps a quarter, can be explored with existing telescopes.[2] The small

[1] A systematic error of the order of 0·35 mag., in the limiting magnitudes of the surveys, as compared with the magnitudes of the nebulae used in the formulation of the law of red-shifts, would be required in order to fully satisfy the relation $C_v = d\lambda/\lambda$.

[2] The volume of this universe would be $2\pi^2R^3$, where R is the radius of curvature, or about 2×10^{27} cubic light-years. The universe would contain about 400 million nebulae.

volume of the universe is another strange and dubious conclusion. The familiar interpretation of red-shifts as velocity-shifts very seriously restricts not only the time-scale, the age of the universe, but the spatial dimensions as well. On the other hand, the alternative possible interpretation, that red-shifts are not velocity-shifts, avoids both difficulties, and presents the observable region as an insignificant sample of a universe that extends indefinitely in space and in time.

The Permissible Type of an Expanding Universe

These conclusions in themselves are important factors in the study of expanding models of the universe, but the implications of the empirical data can be pushed still farther. The nature of the expansion and the nature of the spatial curvature are now determined, but there still remains a third arbitrary element in the general theory, namely, the cosmological constant. This constant occurs in formulae connecting both of the other elements with the contents of the universe. Now we know the nebulae well enough, but we do not know what lies between the nebulae. Consequently, we cannot directly determine the numerical value of the cosmological constant. Nevertheless, with the help of two presumably necessary assumptions, we can make a reliable estimate of the order of the constant. The assumptions are that neither the mean density in the universe, nor the mean pressure, is less than zero—that the universe is not less than empty, either of matter or of radiation. Then, with the aid of the known elements of expansion and curvature, we can set upper and lower limits between which the constant must lie. Fortunately, the limits are so close together that we may state the approximate value of the constant with confidence. It is about $4{\cdot}5 \times 10^{-18}$ (years)$^{-2}$.

The significant features are the facts that the constant

is positive, and is slightly larger than a certain critical value—the value it would have in a particular, unstable model called the Einstein universe. These facts assign the actual world to a single class known as monotonic, or 'ever-expanding', universes. The other elements are sufficient to identify, within the class, a unique model known as the ever-expanding universe of the first kind.

The model expands without reversal. The radius increases from zero to infinity. Past time is finite, future time is infinite. Comparatively recently, perhaps a thousand million years ago, the model started to expand from a small compact mass. The expansion was very rapid at first, but it has steadily slowed down to the rate we measure to-day. At each moment the model is homogeneous, the contents are uniformly distributed, but as time goes on the mean density diminishes, the average distance between neighbouring nebulae increases. Eventually a state of complete isolation will be reached.

The disturbing features of this picture of the universe is the very small scale both in time and in space. With existing telescopes we explore at least a third of the 'age' and about a quarter of the volume. And there is a third questionable feature. The general formulae indicate a definite relation between spatial curvature, the cosmological constant, and the contents of space. The universe suggested by the surveys is very massive. The mean density is of the order of a thousand times that which can be accounted for by the nebulae we observe. The excess matter, if it exists, must be scattered between the luminous nebulae.

We suppose that such internebular material would probably be in the form of dust or ionized gas. In that case the medium would absorb light, and would be readily detected. But we find no trace of space absorption; space is sensibly transparent. Therefore, the

matter, if it exists, must be in some unlikely form such as chunks or non-ionized gas. Even then, the necessary quantity is barely permissible. For, on other evidence, we can set an upper limit to the possible density, regardless of the form of the material. We find no trace of internebular material, but our investigations have been pushed only to a certain limit. At the moment, that limit is just about the density corresponding to the curvature. Thus the theory might be valid provided the universe were packed with matter to the very threshold of perception. Nevertheless, the ever-expanding model of the first kind seems rather dubious. It cannot be ruled out by the observations, but it suggests a forced interpretation of the data.

The disturbing features are all introduced by the recession factors, by the assumption that red-shifts are velocity-shifts. The departure from a linear law of red-shifts, the departure from uniform distribution, the curvature necessary to restore homogeneity, the excess material demanded by the curvature, each of these is merely the recession factor in another form. These elements identify a unique model among the array of possible expanding worlds, and, in this model, the restriction in the time-scale, the limitation of the spatial dimensions, the amount of unobserved material, is each equivalent to the recession factor.

On the other hand, if the recession factor is dropped, if red-shifts are not primarily velocity-shifts, the picture is simple and plausible. There is no evidence of expansion and no restriction of the time-scale, no trace of spatial curvature, and no limitation of spatial dimensions. Moreover, there is no problem of internebular material. The observable region is thoroughly homogeneous; it is too small a sample to indicate the nature of the universe at large. The universe might even be an expanding model, provided the rate of

expansion, which pure theory does not specify, is inappreciable. For that matter, the universe might even be contracting.

Conclusion

In conclusion, we may recall the purpose of the discussion, the precise formulations of the laws of red-shifts and of nebular distribution, using two different scales of distances corresponding to alternative, possible interpretations of red-shifts. It was hoped that the wrong scale, the wrong interpretation, might lead to anomalies or discrepancies which could be detected or at least suspected. The hope is only partially realized. Suspicions are aroused, but they have not been fully substantiated. The observations now available furnish a consistent picture on either scale.

The facts of the case may be presented in a brief table where the laws are expressed both with symbols and with numerical terms. The table also includes the elements derived from the recession factor which identify our universe among the possible worlds offered by relativistic cosmology. The unit of time is the year, and the unit of distance, the light-year, except that the mean density is expressed in grammes per cubic centimetre.

DESCRIPTION OF PLATE VII

The photograph shows a small region of a typical field in the deepest survey using two-hour exposures with the 100-inch reflector. The faintest nebulae identified with confidence are about the 21st (photographic) magnitude, or a million times fainter than the faintest naked-eye stars. These vague spots are the images of great stellar systems, averaging about 210 million times as bright as the sun. They appear so faint because their average distance is of the order of 400 million light-years. The survey records, on the average, about 38 nebulae in each area of the sky equal to that of the plate (about 2/15 the area of the full moon).

As the depth of the explorations increase, the stars in the stellar system tend to thin out while the recorded nebulae steadily accumulate. About half a magnitude beyond the limit of the survey, and in the directions of the galactic poles, the nebulae should be as numerous as the stars. This level of equality has actually been reached on a few photographs made under exceptional conditions with the 100-inch reflector.

The bright star, surrounded by a halation circle (light reflected from the back of the plate), is B.D. $+35^\circ$ 2731. Nebulae are indicated by arrows.

PLATE VII

A SAMPLE OF THE UNIVERSE

SUMMARY OF OBSERVATIONAL RESULTS

(a) *Laws of Red-Shifts*

(Stationary Universe) $d\lambda/\lambda = kr$
(Expanding ,,) $= kr + lr^2 + \ldots$
($k = 5{\cdot}37 \times 10^{-10}$; $l = 2{\cdot}54 \times 10^{-19}$; r = distance)

(b) *Laws of Nebular Distribution*

(Stationary Universe) $\log_{10} N_m = 0{\cdot}6 m_c + C_s$
(Expanding ,,) $= 0{\cdot}6(m_c - d\lambda/\lambda + C_v) + C_e$

$C_s = -4{\cdot}437$; $C_e = -4{\cdot}391$; N_m = number of nebulae over entire sky, brighter than m_c; m_c = apparent magnitude corrected for local obscuration and for energy effects of red-shifts; $d\lambda/\lambda$ = mean red-shift at distance corresponding to m_c. The term C_v is represented by the equation

$$0{\cdot}6 C_v = \log_{10} \left[(3/2x^3)\{\sin^{-1}x - x\sqrt{(1-x^2)}\}\right],$$

in which $x = r/R_0$, where R_0 is the present radius of spatial curvature, and r is the distance corresponding to apparent magnitude $(m_c - d\lambda/\lambda)$.

(c) *Derived Constants in an Expanding Universe*

Present Radius of Spatial Curvature $= R_0 = 4{\cdot}7 \times 10^8$ (light-years).
Cosmological Constant $= 4{\cdot}5 \times 10^{-18}$ (years)$^{-2}$.
Present Mean Density $= 6 \times 10^{-27}$ gm./cm.^{3}

The table neatly summarizes the explorations in the observable region of space, the immense region within which reliable data can be gathered with the largest telescope in operation. The results are a definite step in the observational approach to cosmology. The field is entered, the problems formulated, and the solutions restricted to a fairly narrow range. But the essential clue, the interpretation of red-shifts, must still be unravelled. The former sense of certainty has faded and the clue stands forth as a problem for investigation.

Larger telescopes may resolve the question, or theory may be revised to account for the new data. But with regard to relativistic cosmology in its present form, and the observations now available, the conclusion can be

stated quite simply. Two pictures of the universe are sharply drawn. Observations, at the moment, seem to favour one picture, but they do not rule out the other. We seem to face, as once before in the days of Copernicus, a choice between a small, finite universe, and a universe indefinitely large plus a new principle of nature.

INDEX

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRESS, OXFORD
BY JOHN JOHNSON, PRINTER TO THE UNIVERSITY